

Appendix

i). Estimation of decomposed factors \mathbf{U} and \mathbf{V}

Given the distributions of $p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda)$, $p(\mathbf{U}|\gamma)$, $p(\mathbf{V}|\gamma)$ and $p(\gamma)$. We can use Variational Bayes approach to get the approximation to the posterior distributions of \mathbf{U} and \mathbf{V} , which can be decomposed as independent distributions between their rows. The posterior distribution of the i^{th} row $\mathbf{u}_{i\cdot}$ of \mathbf{U} , the derivation can be expressed as

$$\begin{aligned}\ln q(\mathbf{U}) &= \langle \ln p(\mathbf{Y}', \lambda, \mathbf{z}) \rangle_{\mathbf{z}|\mathbf{U}} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta, \lambda)) \rangle_{\mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) p(\mathbf{U}|\gamma)) \rangle_{\mathbf{V}, \mathbf{S}, \gamma, \beta} + \text{const} \\ &= \left\langle \sum_i \sum_j -\frac{1}{2} \beta (y'_{ij} - \mathbf{u}_{i\cdot} \mathbf{v}_j^T - \lambda s_{ij})^2 \right. \\ &\quad \left. + \sum_j (-\frac{1}{2} \mathbf{u}_{i\cdot}^T (\gamma_j^{-1} \mathbf{E}_K)^{-1} \mathbf{u}_{\cdot j}) \right\rangle_{\mathbf{V}, \mathbf{S}, \gamma, \beta} + \text{const} \\ &= \sum_i -\frac{1}{2} (\mathbf{u}_{i\cdot} (\langle \beta \rangle \langle \mathbf{V}^T \mathbf{V} \rangle + \Gamma) \mathbf{u}_{i\cdot}^T \\ &\quad - 2 \mathbf{u}_{i\cdot} \langle \beta \rangle \langle \mathbf{V} \rangle^T (\mathbf{y}'_{i\cdot} - \lambda \mathbf{s}_{i\cdot})^T) + \text{const} \\ q(\mathbf{u}_{i\cdot}) &\propto \exp(-\frac{1}{2} (\mathbf{u}_{i\cdot} (\langle \beta \rangle \langle \mathbf{V}^T \mathbf{V} \rangle + \Gamma) \mathbf{u}_{i\cdot}^T \\ &\quad - 2 \mathbf{u}_{i\cdot} \langle \beta \rangle \langle \mathbf{V} \rangle^T (\mathbf{y}'_{i\cdot} - \lambda \mathbf{s}_{i\cdot})^T))\end{aligned}$$

If a row vector \mathbf{a} obey the Gaussian distribution, the mean is \mathbf{u} , the covariance is Σ , it can be expresses by a function as follows

$$p(\mathbf{a}) \propto \exp(-\frac{1}{2} (\mathbf{a} - \mathbf{u}) \Sigma^{-1} (\mathbf{a}^T - \mathbf{u}^T))$$

ii). Estimation of \mathbf{S}

The posterior distribution of \mathbf{S} is found to be decomposed on each coefficient s_{ij} , the derivation can be expressed as:

$$\begin{aligned}\ln q(\mathbf{S}) &= \langle \ln p(\mathbf{Y}', \lambda, \mathbf{z}) \rangle_{\mathbf{z}|\mathbf{S}} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta, \lambda)) \rangle_{\mathbf{U}, \mathbf{V}, \gamma, \mathbf{a}, \beta} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) p(\mathbf{S}|\lambda, \mathbf{a})) \rangle_{\mathbf{U}, \mathbf{V}, \mathbf{a}, \beta} + \text{const} \\ &= \left\langle \sum_i \sum_j -\frac{1}{2} \beta (y'_{ij} - l_{ij} - \lambda s_{ij})^2 + \sum_i \sum_j (-\frac{1}{2} \lambda^q \alpha_{ij} s_{ij}^2) \right\rangle_{\mathbf{U}, \mathbf{V}, \mathbf{a}, \beta} + \text{const} \\ &= \sum_i \sum_j -\frac{1}{2} ((\lambda^2 \langle \beta \rangle + \lambda^q \langle \alpha_{ij} \rangle) s_{ij}^2 - 2 \lambda \langle \beta \rangle (y'_{ij} - l_{ij}) s_{ij}) + \text{const} \\ q(s_{ij}) &\propto \exp(-\frac{1}{2} ((\lambda^2 \langle \beta \rangle + \lambda^q \langle \alpha_{ij} \rangle) s_{ij}^2 - 2 \lambda \langle \beta \rangle (y'_{ij} - l_{ij}) s_{ij}))\end{aligned}$$

iii). Estimation of γ

The posterior density of γ_j can be formulated as:

$$\begin{aligned}\ln q(\gamma) &= \langle \ln p(\mathbf{Y}', \lambda, \mathbf{z}) \rangle_{\mathbf{z}|\gamma} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta, \lambda)) \rangle_{\mathbf{U}, \mathbf{V}, \mathbf{S}, \mathbf{a}, \beta} + \text{const} \\ &= \langle \ln p(\mathbf{U}|\gamma) + \ln p(\mathbf{V}|\gamma) + \ln p(\gamma) \rangle_{\mathbf{U}, \mathbf{V}} + \text{const} \\ &= \left\langle \ln \left(\left| \gamma_j^{-1} \mathbf{E}_K \right|^{-\frac{1}{2}} \exp(-\frac{1}{2} \mathbf{u}_{\cdot j}^T (\gamma_j^{-1} \mathbf{E}_K)^{-1} \mathbf{u}_{\cdot j}) \left| \gamma_j^{-1} \mathbf{E}_N \right|^{-\frac{1}{2}} \right. \right. \\ &\quad \left. \left. \times \exp(-\frac{1}{2} \mathbf{v}_{\cdot j}^T (\gamma_j^{-1} \mathbf{E}_N)^{-1} \mathbf{v}_{\cdot j}) \gamma_j^{a-1} \exp(-b \gamma_j) \right) \right\rangle_{\mathbf{U}, \mathbf{V}} + \text{const} \\ &= \ln(\gamma_j^{a-1+\frac{K+N}{2}} \exp(-\frac{1}{2} \gamma_j (\langle \mathbf{u}_{\cdot j}^T \mathbf{u}_{\cdot j} \rangle + \langle \mathbf{v}_{\cdot j}^T \mathbf{v}_{\cdot j} \rangle + 2b)) + \text{const} \\ q(\gamma_j) &\propto \exp(\gamma_j^{a-1+\frac{K+N}{2}} \exp(-\frac{1}{2} \gamma_j (\langle \mathbf{u}_{\cdot j}^T \mathbf{u}_{\cdot j} \rangle + \langle \mathbf{v}_{\cdot j}^T \mathbf{v}_{\cdot j} \rangle + 2b)))\end{aligned}$$

iv). Estimation of \mathbf{a}

The posterior probability density of α_{ij} is found as a Gamma distribution, the derivation is expressed as follows:

$$\begin{aligned}\ln q(\mathbf{a}) &= \langle \ln p(\mathbf{Y}', \lambda, \mathbf{z}) \rangle_{\mathbf{z}|\mathbf{a}} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta, \lambda)) \rangle_{\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \beta} + \text{const} \\ &= \langle \ln(p(\mathbf{S}|\lambda, \mathbf{a}) p(\mathbf{a})) \rangle_{\mathbf{S}} + \text{const} \\ &= \langle \ln p(\mathbf{S}|\lambda, \mathbf{a}) + \ln p(\mathbf{a}) \rangle_{\mathbf{S}} + \text{const} \\ &= \left\langle \ln \left(\prod_i \prod_j (\alpha_{ij})^{\frac{1}{2}} \exp(-\frac{1}{2} \lambda^q \alpha_{ij} s_{ij}^2) \right) \right. \\ &\quad \left. + \ln \left(\prod_i \prod_j (\alpha_{ij})^{-1} \right) \right\rangle_{\mathbf{S}} + \text{const}\end{aligned}$$

$$\ln q(\alpha_{ij}) = \ln((\alpha_{ij})^{\frac{1}{2}-1} \exp(-\frac{1}{2} \lambda^q \alpha_{ij} \langle s_{ij}^2 \rangle)) + \text{const}$$

$$q(\alpha_{ij}) \propto (\alpha_{ij})^{\frac{1}{2}-1} \exp(-\frac{1}{2} \lambda^q \alpha_{ij} (\langle s_{ij} \rangle^2 + \Sigma_{ij}^{\mathbf{S}}))$$

v). Estimation of noise precision β

The posterior probability density of β is found as a Gamma distribution, the derivation is shown as follows:

$$\begin{aligned}\ln q(\beta) &= \langle \ln p(\mathbf{Y}', \lambda, \mathbf{z}) \rangle_{\mathbf{z}|\beta} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}, \beta, \lambda)) \rangle_{\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \mathbf{a}} + \text{const} \\ &= \langle \ln(p(\mathbf{Y}'|\mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) p(\beta)) \rangle_{\mathbf{U}, \mathbf{V}, \mathbf{S}} + \text{const} \\ q(\beta) &\propto \beta^{\frac{KN}{2}} \beta^{-1} \exp(-\frac{1}{2} \beta \langle \|\mathbf{Y}' - \mathbf{U}\mathbf{V}^T - \lambda \mathbf{S}\|_F^2 \rangle) \\ &\propto \beta^{\frac{KN}{2}-1} \exp(-\frac{1}{2} \beta \langle \|\mathbf{Y}' - \mathbf{U}\mathbf{V}^T - \lambda \mathbf{S}\|_F^2 \rangle)\end{aligned}$$

$$\begin{aligned} \left\langle \left\| \mathbf{Y} - \mathbf{U}\mathbf{V}^T - \lambda \mathbf{S} \right\|_F^2 \right\rangle &= \left\| \mathbf{Y} - \langle \mathbf{U} \rangle \langle \mathbf{V} \rangle^T - \lambda \langle \mathbf{S} \rangle \right\|_F^2 + tr \left(N \langle \mathbf{U} \rangle^T \langle \mathbf{U} \rangle \mathbf{\Sigma}^v \right) \\ &+ tr \left(K \langle \mathbf{V} \rangle^T \langle \mathbf{V} \rangle \mathbf{\Sigma}^u \right) + tr \left(KN \mathbf{\Sigma}^u \mathbf{\Sigma}^v \right) + \lambda^2 \sum_{i=1}^K \sum_{j=1}^N \Sigma_{ij}^s \end{aligned}$$